

# **Supplementary Information**

for

Jung *et al.*, Impact of sequencing depth in ChIP-seq experiments

## **Supplementary Content**

Supplementary Table 1 ----- pp 2

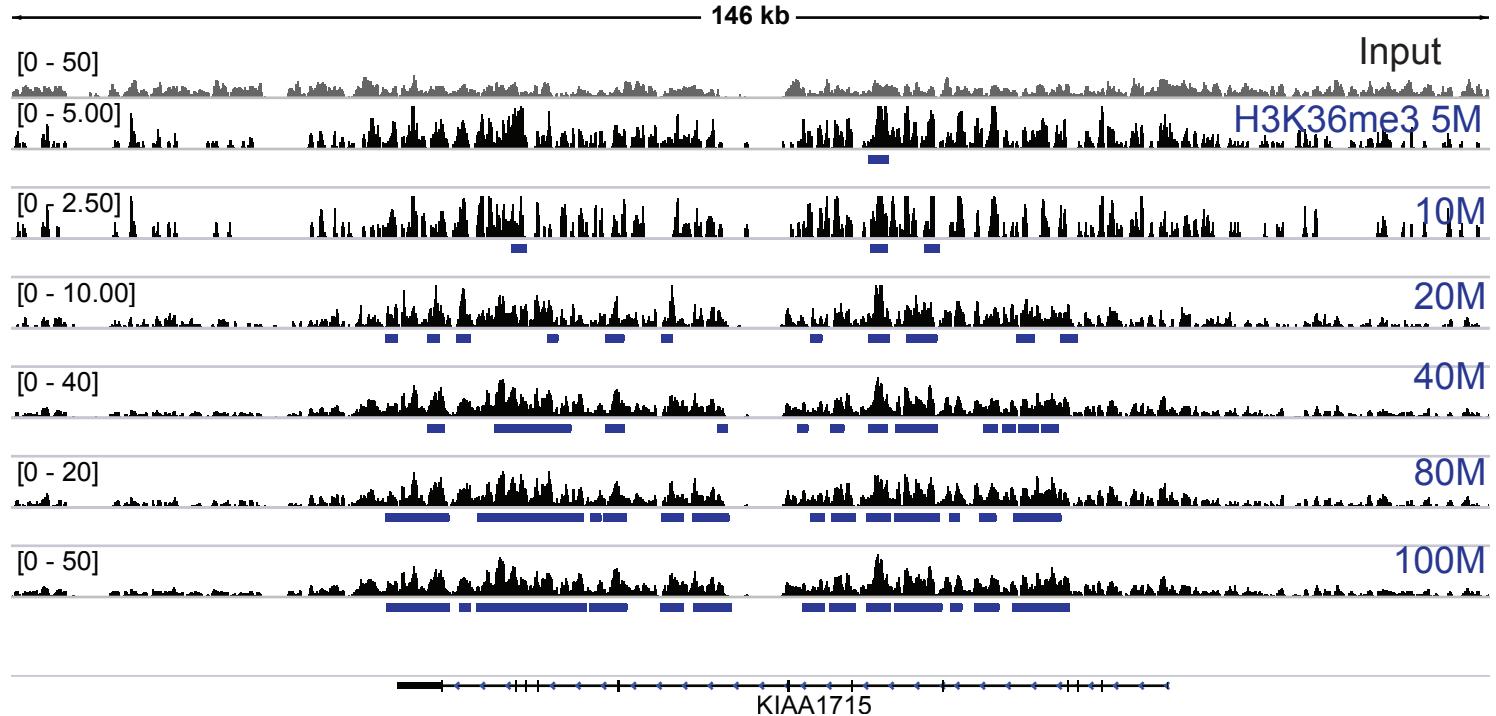
Supplementary Figures 1-10 ----- pp 3-12

**Table S1.** The list of ChIP-seq experiments and their database identifiers.

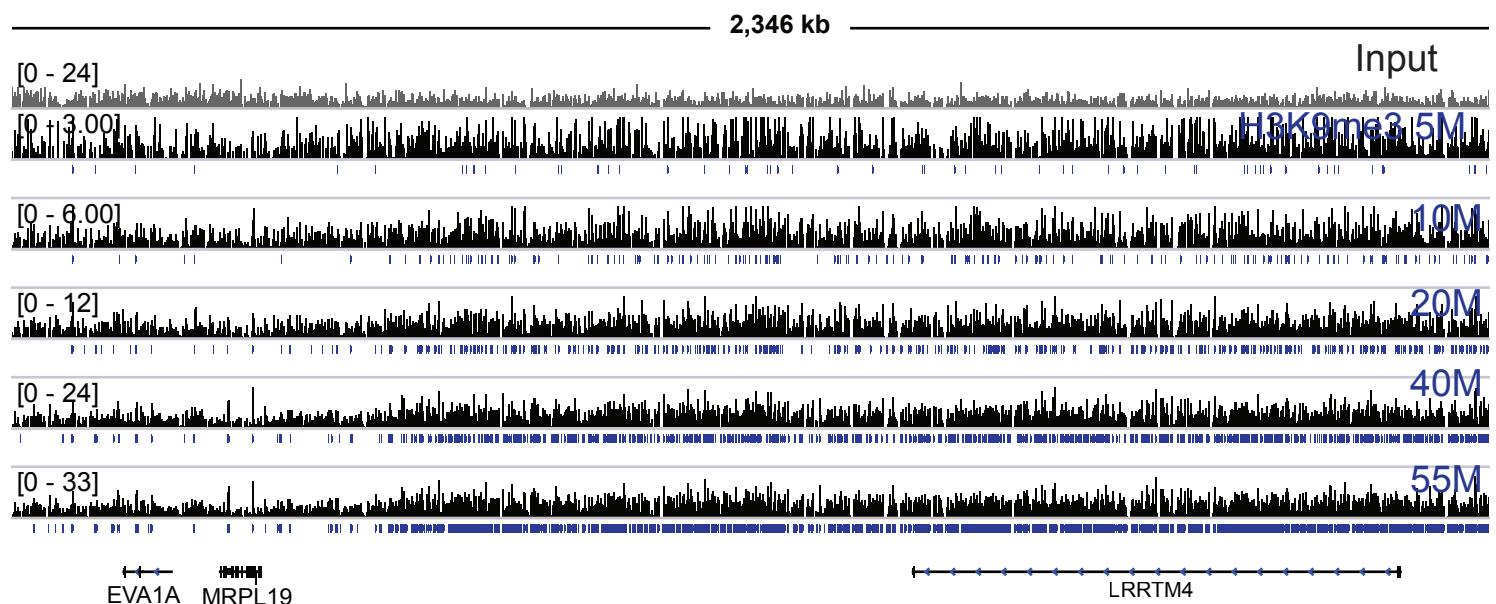
\*: deep-sequencing datasets.

Species	Cell/tissue type	Factor	Accession number
Human *	A549	H3K4me3	SRA091729
Human *	A549	H3K36me3	SRA091729
Human *	A549	H3K27me3	SRA091729
Human *	A549	H3K9me3	SRA091729
Fly *	Late embryos	H3K4me3	SRA091729
Fly *	Late embryos	H3K36me3	SRA091729
Fly *	Late embryos	H3K27me3	SRA091729
Fly *	Late embryos	H3K9me2	SRA091729
Fly	Late embryos	H3K4me1	GSE47281
Fly	Late embryos	H3K36me2	GSE47335
Fly	Late embryos	H3K79me1	GSE47239
Fly	Late embryos	H3K79me2	GSE47252
Fly	Late embryos	H4K20me1	GSE47293
Fly	Late embryos	H2AV	GSE47268
Fly	Late embryos	H3K27ac	GSE47237
Fly	Late embryos	H4K16ac	GSE47245
Fly	Late embryos	H3K27me2	GSE47277
Fly	Late embryos	H4	GSE47291
Fly	Late embryos	H3K36me1	GSE47241
Fly	Late embryos	H3K18ac	GSE47274
Fly	Late embryos	H3K9me1	GSE47240
Fly	Late embryos	H3K9acS10P	GSE47287
Fly	Late embryos	H3K23ac	GSE47275
Fly	Late embryos	H1	GSE47309

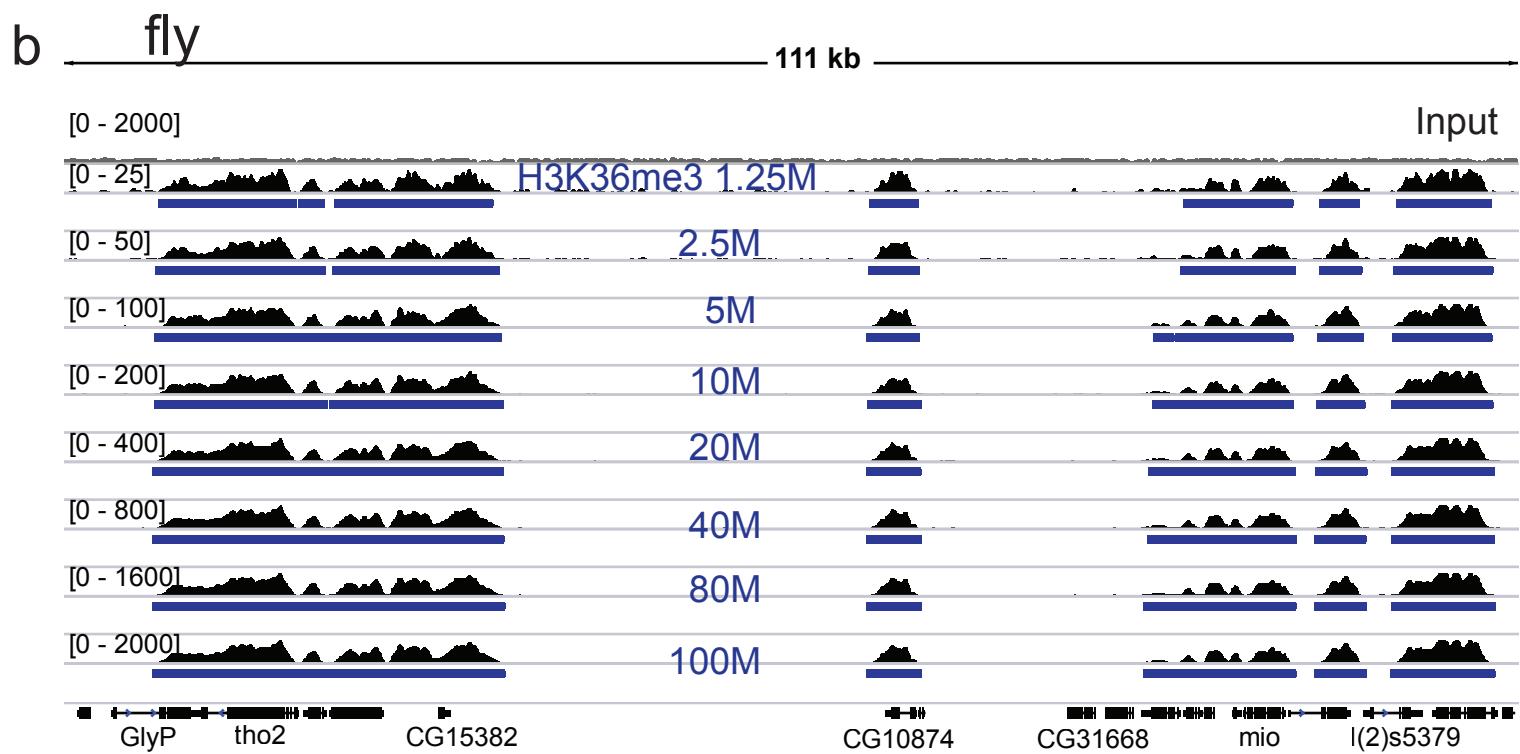
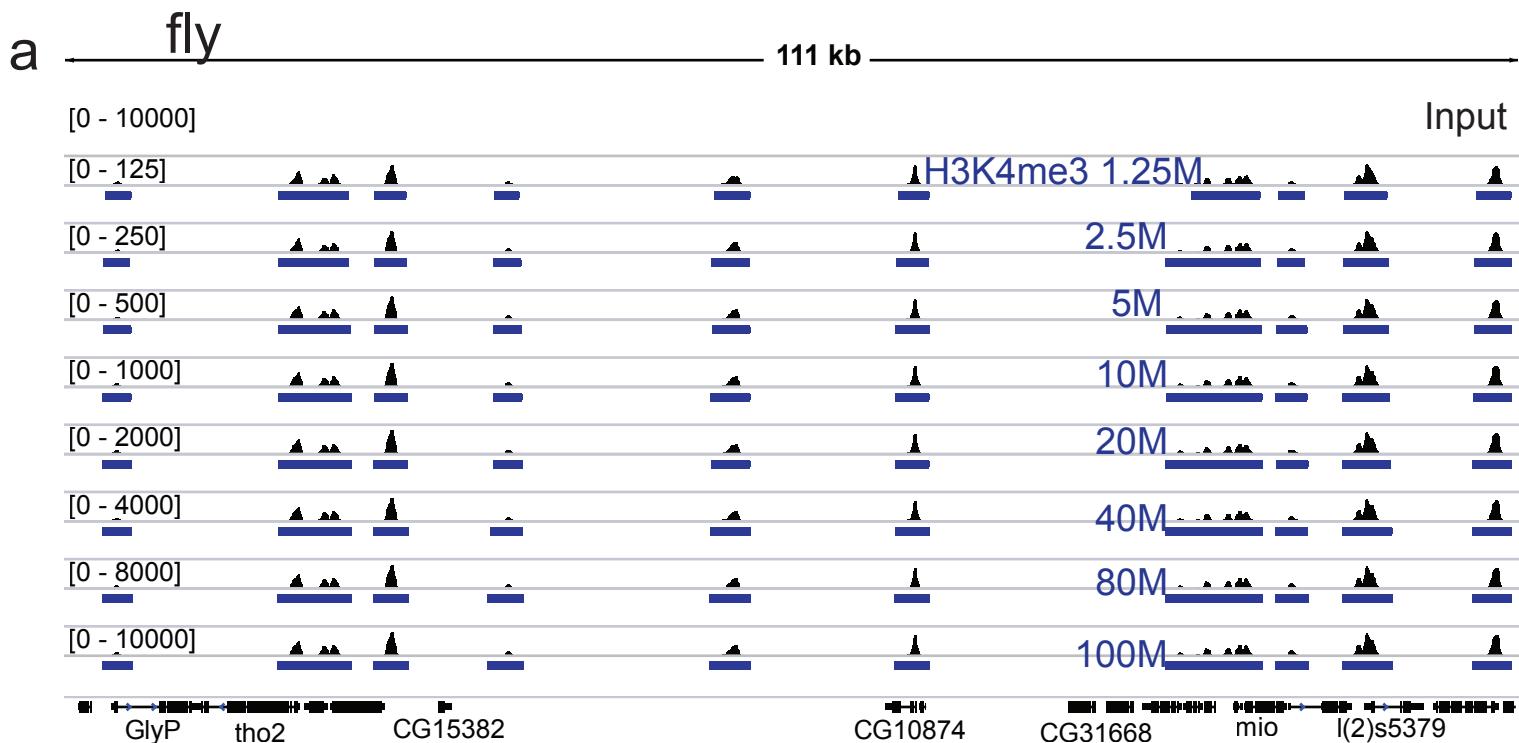
a human



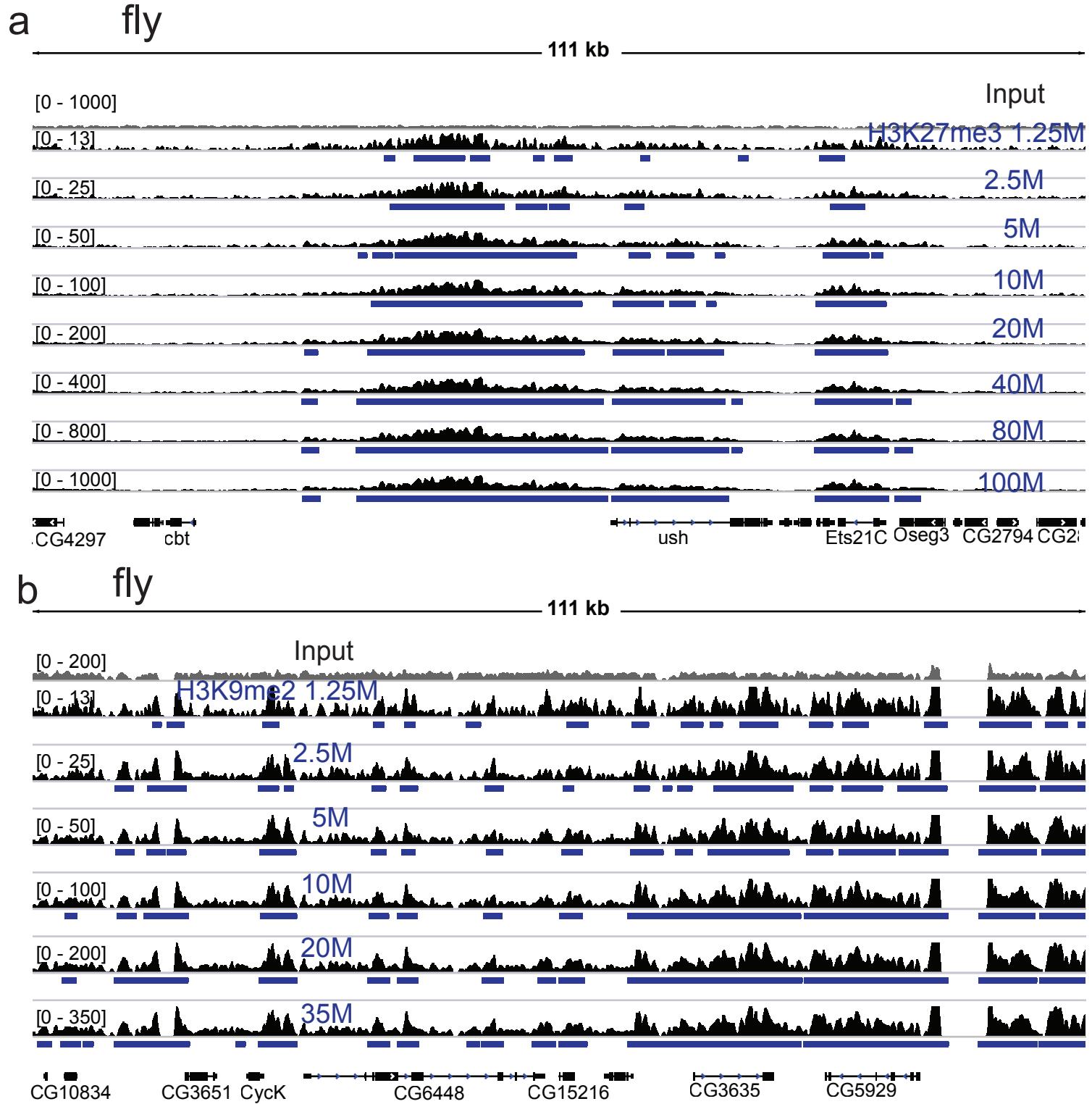
b human



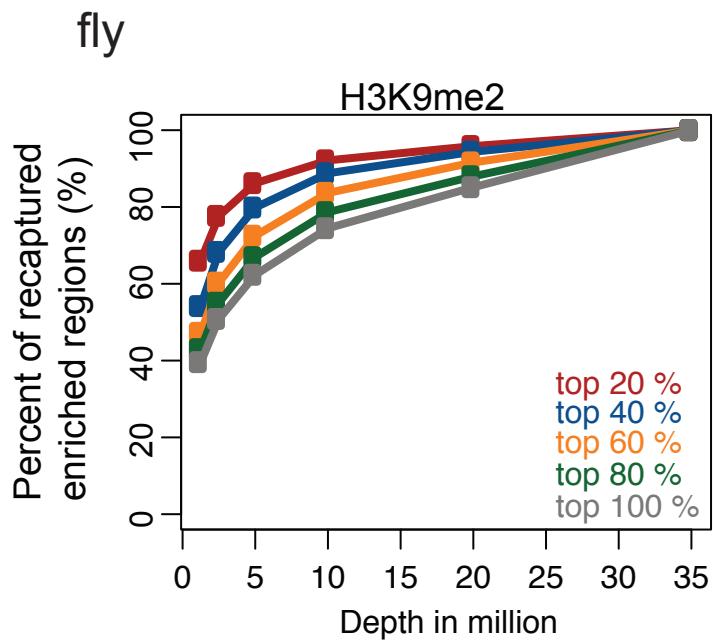
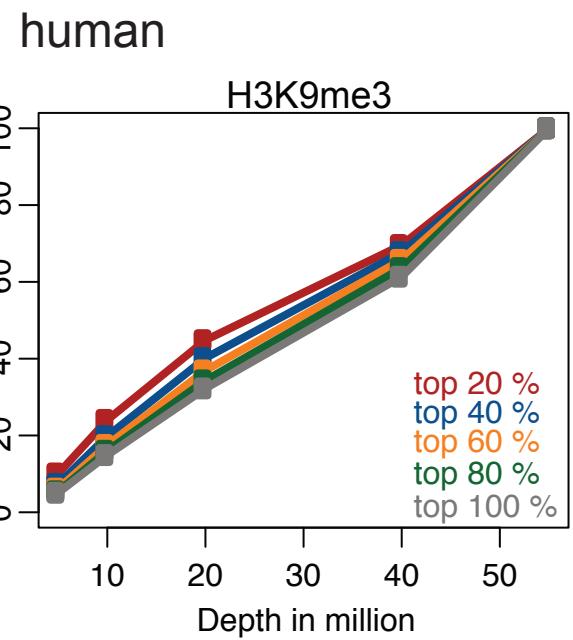
**fig. S1. ChIP tag density profiles for human H3K36me3 and H3K9me3.** **a**, Profiles at different sequencing depths (increase in reads from top to bottom) for H3K36me3 in human A549. An input profile is on the top row for comparison. The blue boxes below the ChIP profiles indicate the significantly enriched regions based on SPP. **b**, ChIP tag density profiles for H3K9me3 in human A549.



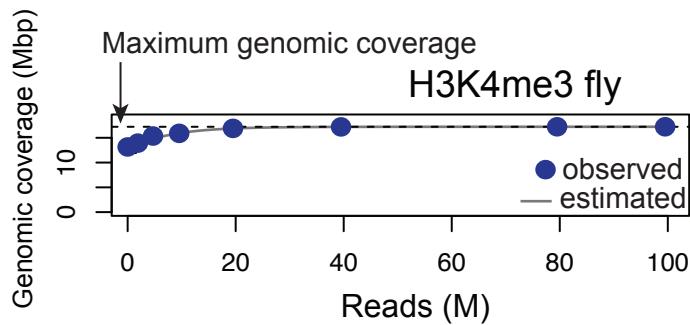
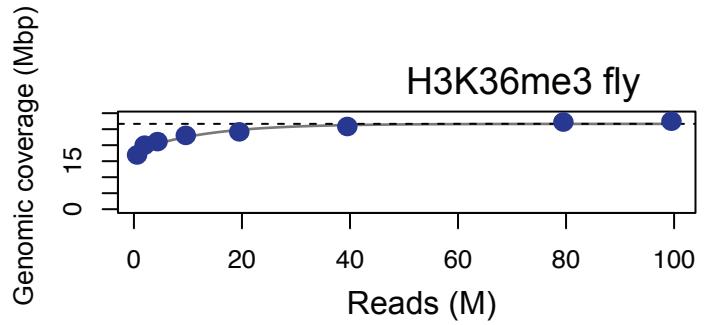
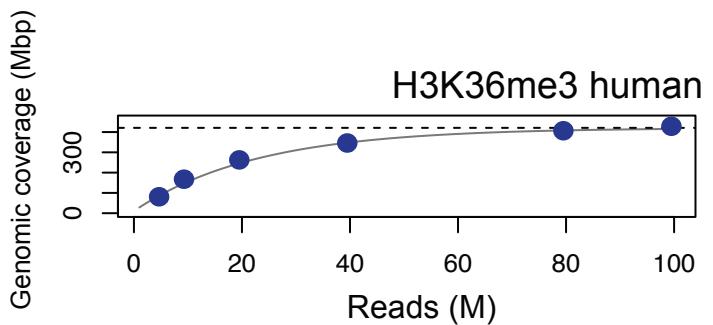
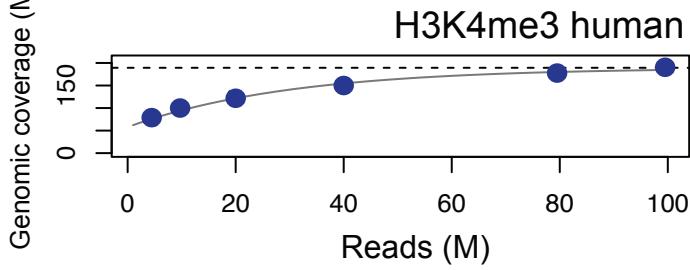
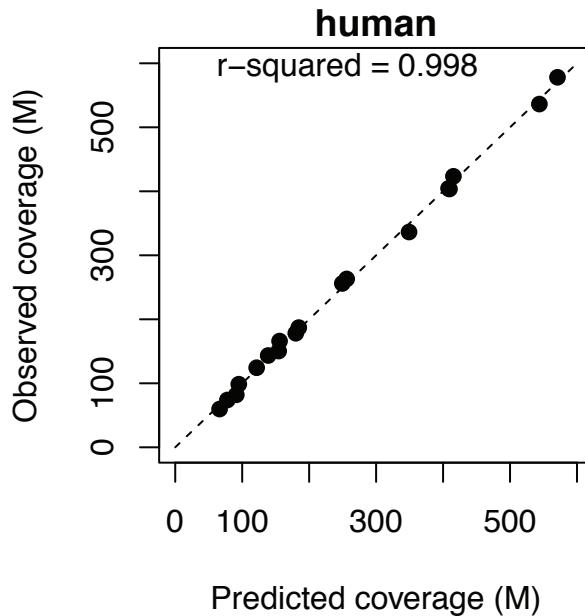
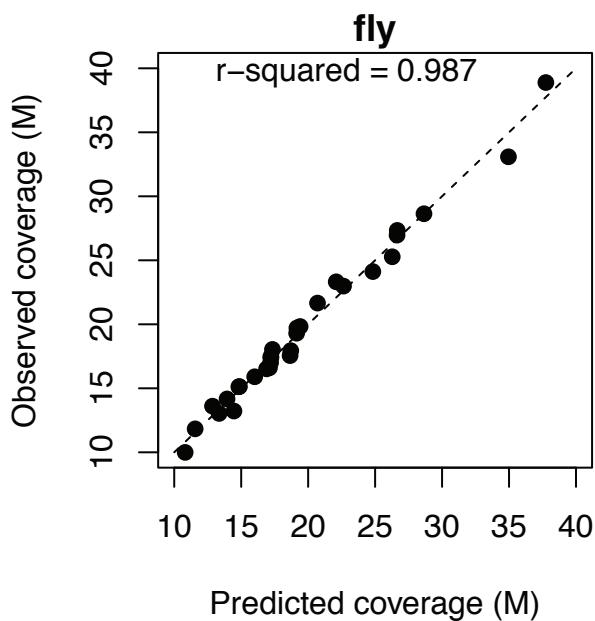
**fig. S2. ChIP tag density profiles for fly H3K4me3 and H3K36me3.** **a**, Profiles at different sequencing depths (increase in reads from top to bottom) for H3K4me3 in fly late embryos. An input profile is on the top row for comparison. The blue boxes below the ChIP profiles indicate the significantly enriched regions based on SPP. **b**, ChIP tag density profiles for H3K36me3 in fly late embryos.



**fig. S3. ChIP tag density profiles for fly H3K27me3 and H3K9me2.** **a**, Profiles at different sequencing depths (increase in reads from top to bottom) for H3K27me3 in fly late embryos. An input profile is on the top row for comparison. The blue boxes below the ChIP profiles indicate the significantly enriched regions based on SPP. **b**, ChIP tag density profiles for H3K9me2 in fly late embryos.

**a****b**

**fig. S4. Percentage of recaptured enrichment regions.** **a**, Percentage of significantly enriched regions from the data of the 35 million reads recaptured in each subsampling of the data for fly H3K9me2. Each line represents the percent of the enriched regions recaptured in each subsampling, for top 20%, 40%, 60%, 80% and all enriched regions from the 35 million reads data. **b**, Percentage of significantly enriched regions from the data of the 55 million reads recaptured in each subsampling of the data for human H3K9me3.

**a****b****c****c**

**fig. S5. Estimated genomic coverage and observed genomic coverage.** **a**, Observed and estimated maximum genomic coverage (at complete saturation) for H3K4me3 in fly (upper) and human (lower). Observed points: blue dots, Estimated values based a simple Poisson model: gray solid lines and Estimated maximum genomic coverage: dashed lines. **b**, Observed and estimated maximum genomic coverage (at complete saturation) for H3K36me3 in fly (upper) and human (lower). **c**, Observed and estimated genomic coverage at each subsampling points from H3K4me3, H3K36me3, H3K27me3 and H3K9me2 for fly (left) and from H3K4me3, H3K36me3 and H3K27me3 for human (right).

**a**

## Using the model

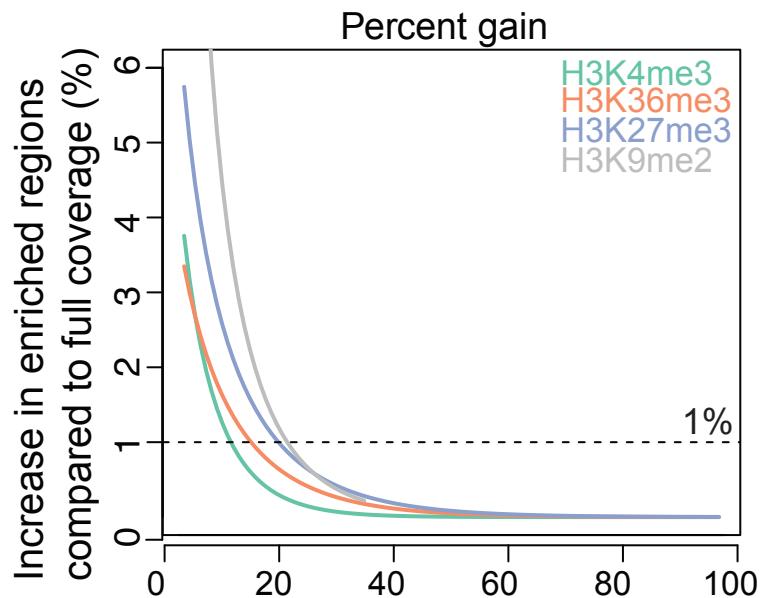
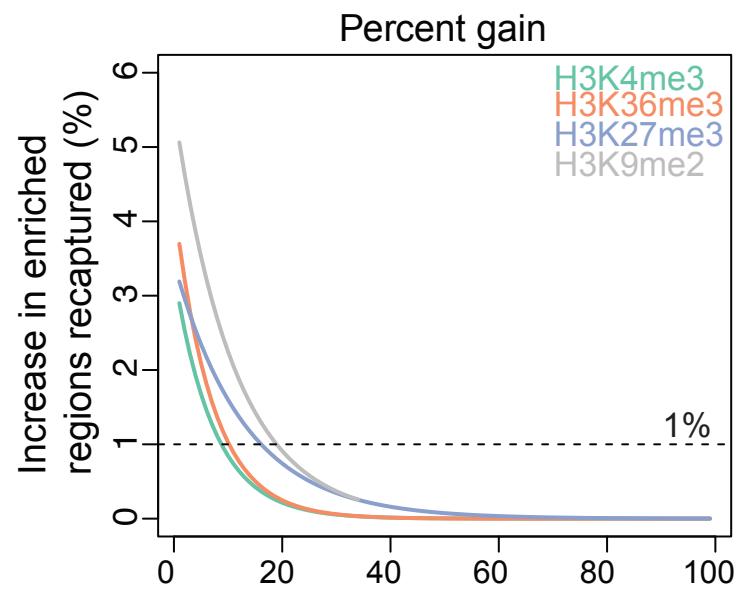
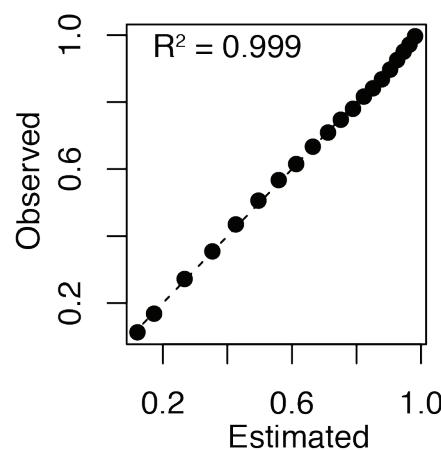
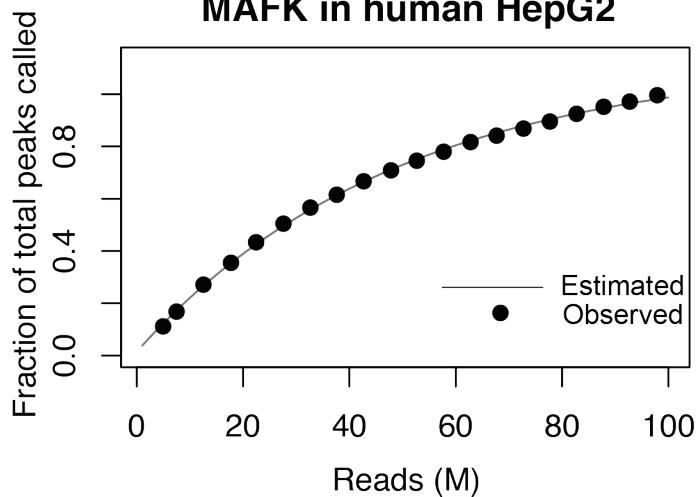


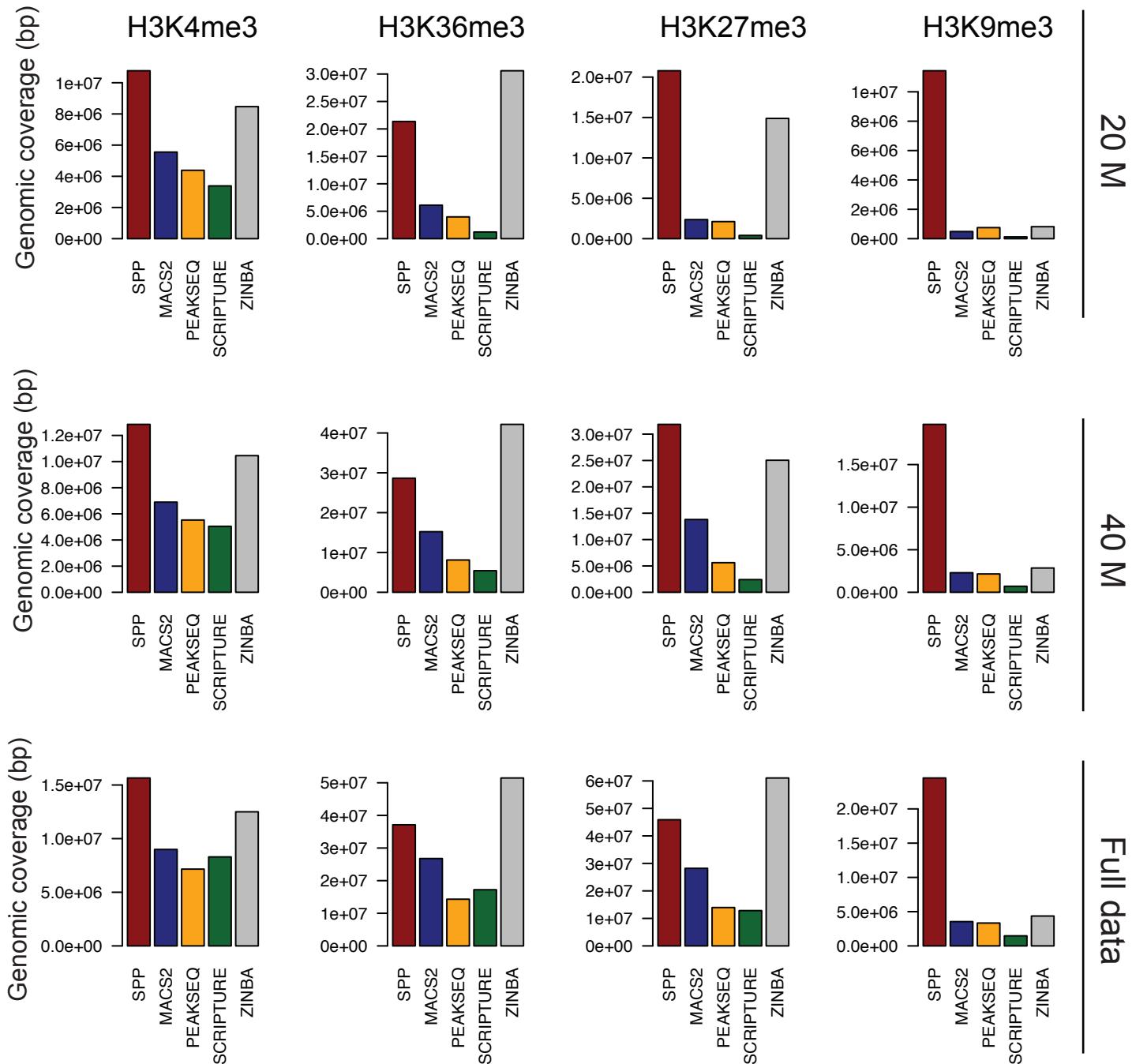
Fig. 2b

**b**

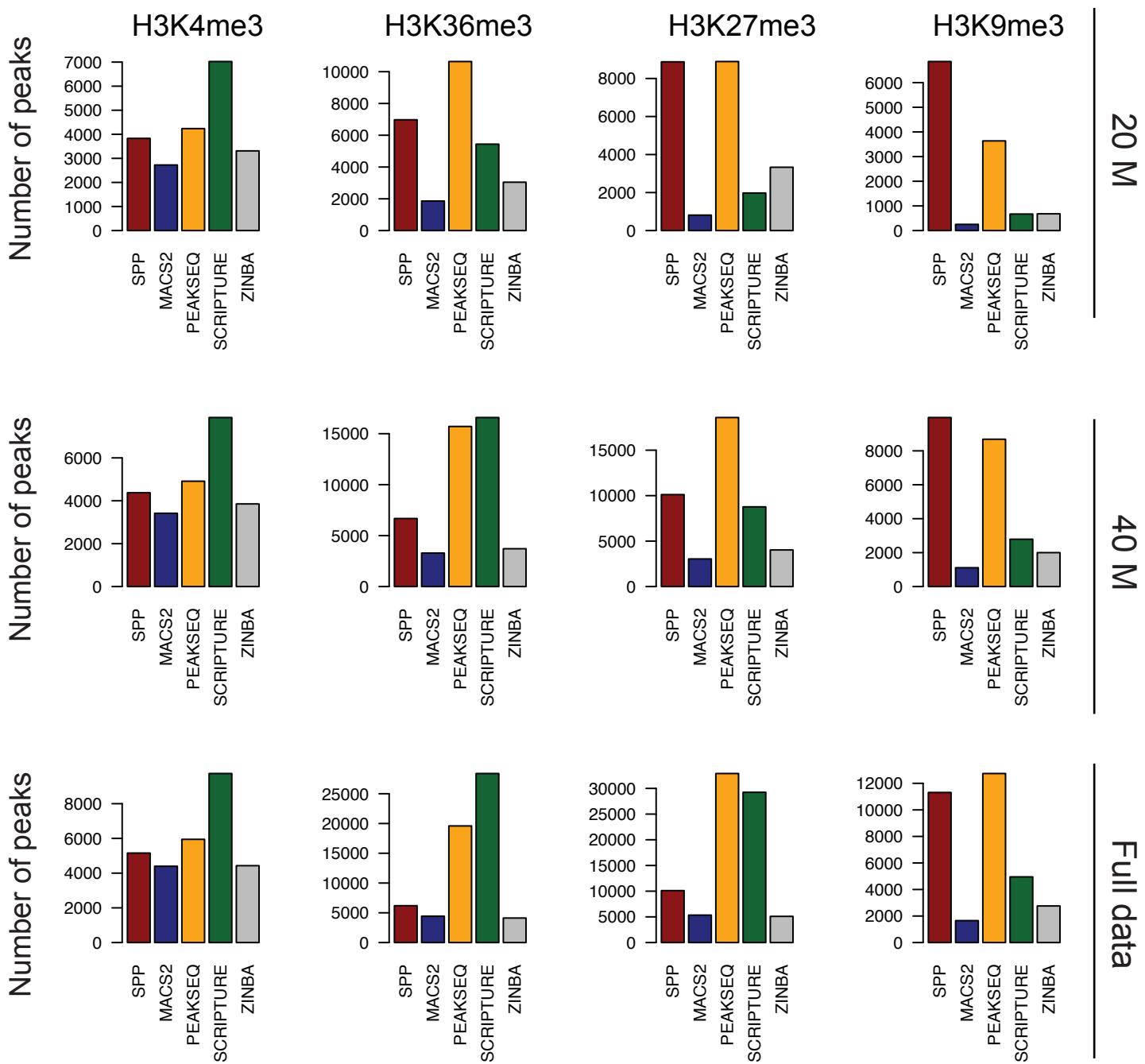
## MAFK in human HepG2



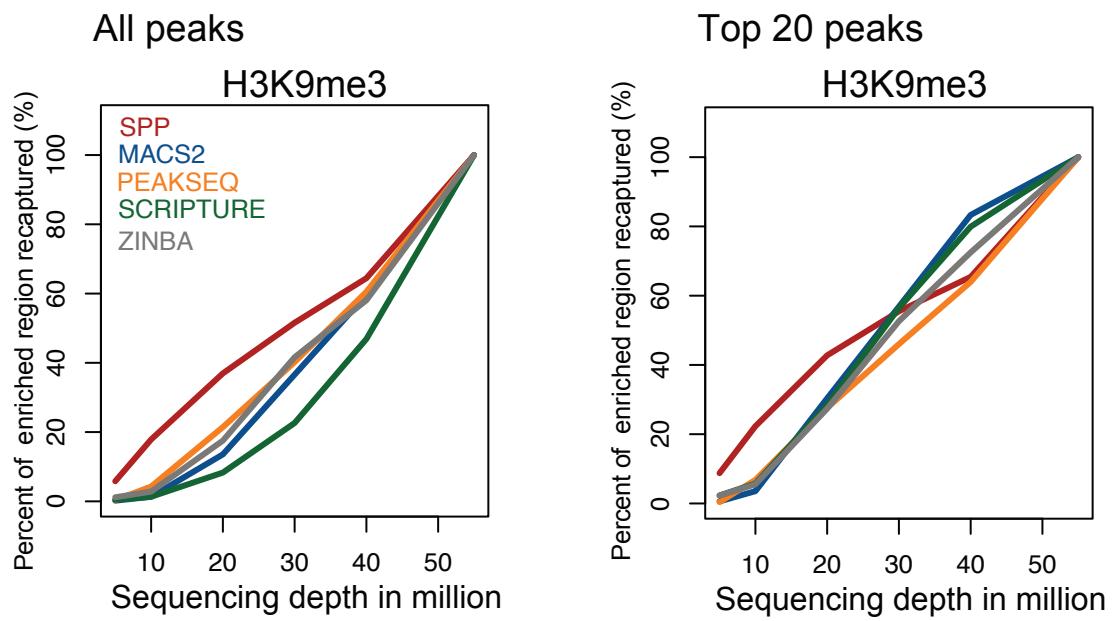
**fig. S6. Comparison between the estimated data and the observed data.** **a**, Estimated sufficient depth. Left: Percentage of increase in the enriched regions estimated compared to the genomic coverage at full saturation using the model. Sufficient depth is the sequencing depth at which the percent gain per 1 million additional sequence reads falls below 1%. Right: the same as in Fig. 2b for comparison. Each line shows percentage of increase in enriched regions recaptured when an additional 1 million reads are sequenced for fly data. **b**, Model fitting to TF data. Left: fraction of peaks called recaptured at subsampling data observed in MAFK in human HepG2 (data reported in Fig. 3b of Landt *et al.* (5)) and the estimated values using the model. For the sample, the observed fraction of peaks recaptured at subsampling reads was well fit to our formula.



**fig. S7. Genomic coverage detected by various algorithms.** **a**, The genomic coverage of detected enriched regions by SPP(red), MACS2 (blue), PeakSeq (yellow), Scripture (green) and Zinba (gray) for human H3K4me3, H3K36me3, H3K27me3 and H3K9me3. Top row: 20 million reads, middle: 40 million reads and bottom: full data (100 million reads for H3K4me3, H3K36me3 and H3K27me3; 55 million reads for H3K9me3). The results from chromosome 1.

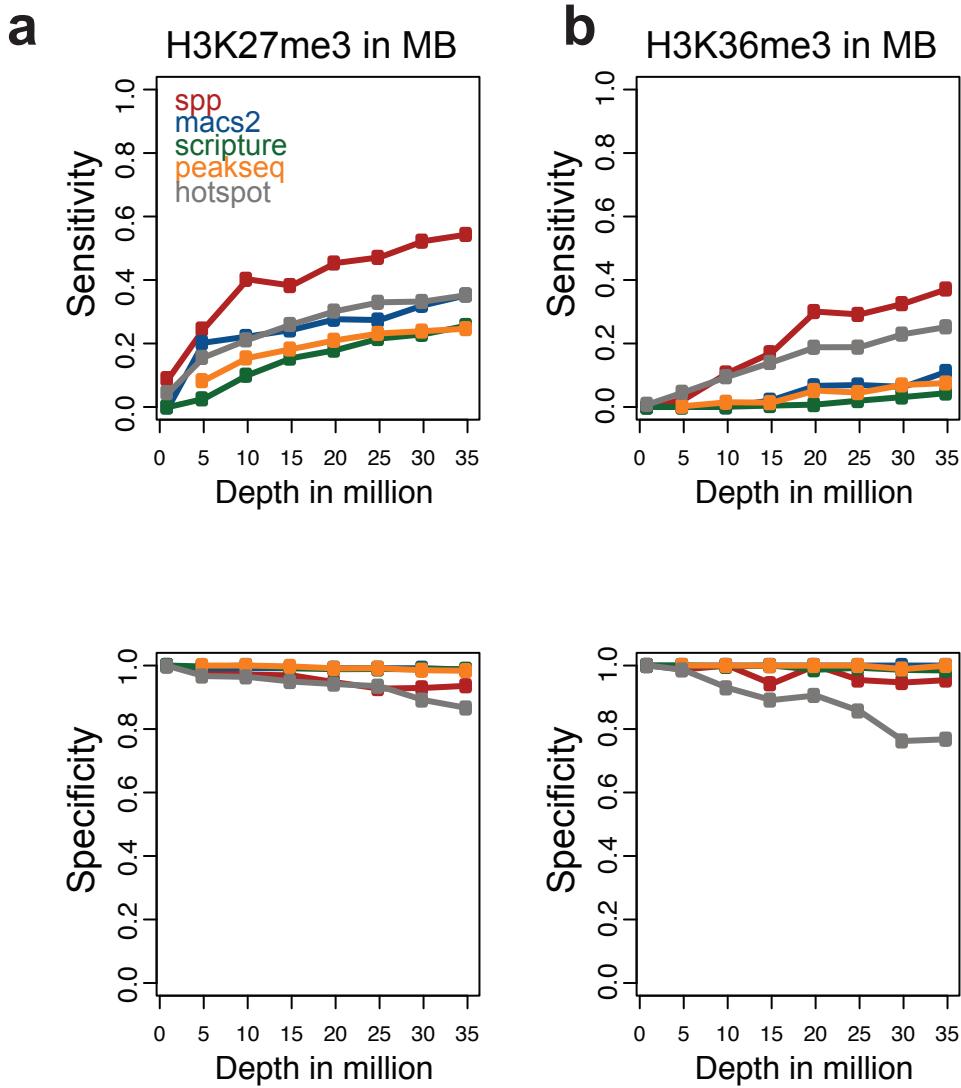


**fig. S8. Number of enriched regions detected by various algorithms.** **a**, Number of enriched regions detected by SPP(red), MACS2 (blue), PeakSeq (yellow), Scripture (green) and Zinba (gray) for human H3K4me3, H3K36me3, H3K27me3 and H3K9me3. Top row: 20 million reads, middle: 40 million reads and bottom: full data (100 million reads for H3K4me3, H3K36me3 and H3K27me3; 55 million reads for H3K9me3). The results from chromosome 1.



**fig. S9. Percentage of enriched regions recaptured by each method for human H3K9me3.** Percentage of all enriched regions identified by each method from the data of the 55 million reads recaptured in different sequencing depths (left). Percentage of top 20% enriched regions identified by each method from the data of the 55 million reads recaptured in different sequencing depths (right).

## mouse



**fig. S10. Comparison of performance of peak calling algorithms using the qPCR validated loci.** **a**, Sensitivity (upper) and specificity (lower) for mouse H3K27me3 in mouse muscle cells MB (myoblasts) (13), considering qPCR validated sites as a true set (14) with various sequencing depths. Sensitivity and specificity were calculated based on the overlapped regions in bp. **b**, Sensitivity (upper) and specificity (lower) for mouse H3K36me3 in mouse muscle cells MB (myoblasts) (13), considering qPCR validated sites as a true set (14) with various sequencing depths.